C. Statement of Need

C.1 Background and Purpose

The Geophysical Fluid Dynamics Laboratory (GFDL), located on Princeton University's Forrestal Campus, Plainsboro, NJ, is a federal research laboratory in the Office of Oceanic and Atmospheric Research (OAR) of the National Oceanic and Atmospheric Administration (NOAA) of the U.S. Department of Commerce (DOC). GFDL performs comprehensive, long lead time research that is fundamental to the mission of NOAA. The goal of this research is to expand the scientific understanding of the physical processes that govern the behavior of the atmosphere and the oceans as complex fluid systems. These systems can then be modeled mathematically and their phenomenology can be studied by complex computer simulations. To help achieve this goal, GFDL proposes to enhance its computing capabilities by acquiring a balanced high-performance computing system (HPCS). The HPCS must provide a complete high-performance computing environment, including very large scalable computing and archiving capacity, and analysis, visualization, networking, and telecommunications capabilities. This enhanced computational capability will replace GFDL's current T932 and T3E supercomputers leased from SGI/Cray Research and will provide all of the facilities needed by GFDL to carry out its computational research.

Increased computational power is essential for GFDL to make progress in solving some of the very difficult problems confronting the climate and weather research community and to support on-going and developing research collaborations within NOAA and with other government agencies. The sharp increase in computing, archival, and analysis capabilities provided by the HPCS acquisition will allow NOAA to most effectively use the scientific talent at GFDL to attack some of the critical problems that currently inhibit progress in decadal-centennial climate projections, seasonal-interannual climate predictions, and the development of the next-generation hurricane prediction system.

This procurement has clearly defined goals from the NOAA IT^2 budget initiative document, entitled "Attacking Frontier Challenges in Climate and Weather Modeling". The research drivers that are the focus of this initiative involve the very leading edge of research activities in climate and weather and address the following four scientific objectives:

- Develop a more realistic model representation of cloud-radiative feedback. This will result in improved regional projections of climate change that will help the broader research community determine the impacts of this change as part of the U.S. climate research program and the IPCC climate change assessment in 2005.

- Identify and evaluate sources of climate "drift" in long-running, higher resolution, coupled climate models, including investigation of the effects of the deep ocean circulation on model behavior.

- Develop the next-generation GFDL coupled research model using more realistic physics, higher resolution, and full ocean-atmosphere-soil coupling and evaluate its skill for seasonal-interannual climate prediction and its capability for elucidating the processes controlling El-Niño-Southern-Oscillation events.

- Develop a more advanced GFDL Hurricane Prediction System to further improve track forecasting accuracy and to provide improved prediction of wind and precipitation fields, storm surge, and changes in storm intensity.

As discussed in the budget initiative document, each of these areas provides potentially enormous benefits to the Nation in terms of reduced costs resulting from improved hurricane predictions, value to agriculture arising from improved seasonal forecasts and interannual climate predictions, and critical value for the Nation through a reduction in uncertainty about global climate change.

In order to fulfill the objective of acquiring the HPCS in FY2000, GFDL will utilize the Department of Commerce's re-engineered acquisition process referred to as Concept of Operations or CONOPS, described in "Department of Commerce Acquisition Process Case for Change" (http://oamweb.osec.doc.gov/conops). The intent of the new process is to create an improved acquisition environment that will benefit the contractor and the government. In order to successfully implement this new process within this acquisition, the government seeks the cooperation of the vendor community in an effort to conduct business in an atmosphere of integrity, openness, and fairness. It is essential that GFDL acquire the best HPCS available for the budgeted level of funding and do so in an expedient and fair manner.

This Statement of Need establishes the purpose, objectives, and the expected results for the HPCS procurement. The contract to be awarded will provide for multiple computer systems and ancillary equipment. The contract will also provide for maintenance and support services.

Definitions of terms used in this Statement of Need may be found in section C.6.

C.2 Procurement Objective

The primary objective of this acquisition is to acquire balanced, comprehensive computing capabilities for GFDL. This includes not only high-performance computing but also corresponding capabilities for data management and archiving, analysis and visualization of model results, and networking and telecommunications that the Laboratory needs to most effectively advance its research programs.

The key to effective use of the HPCS is balance. Each component of the HPCS plays a critical role in maintaining the flow of information through GFDL's model simulations, analyses, visualizations, and ultimately into scientific insight and the dissemination of knowledge to the research community and the other customers of the Laboratory's research. Consequently, the functional capability of the large-scale computers, hierarchical storage management system, analysis and visualization platforms, desktop workstations, and network bandwidth must be well-matched in a way that minimizes bottlenecks to the flow of information while maximizing performance. Achieving this proportionality in the acquired capabilities is an essential goal of this procurement.

Additionally, the computational resources available to GFDL must balance its scientific needs throughout the life of the contract, so GFDL requires a phased delivery of all components of the HPCS. The initial delivery of the HPCS must provide a substantial increase over current capabilities in computational throughput for the Laboratory. At least one substantial upgrade to the sustained throughput must be offered during the base contract period, with archiving and other HPCS capabilities increasing commensurately. The schedule for delivering the upgrade(s) shall be determined by discussions between the contractor and the government.

The period covered under the Project Agreement, FY2000-FY2006, will be divided into a base contract period (FY2000-2003), followed by an option period (FY2004-2006). During the base contract period, the contract will be renewed each year subject to the availability of funds. The decision to exercise the option in FY2004 will be made by evaluating a proposal, submitted by the incumbent contractor at the end of FY2002, that provides detailed specifications for the HPCS during the option period. The contract will be renewed each year subject to the availability of funds. The final year of HPCS operations (FY2006) will overlap with the acquisition and acceptance of successor computational capabilities.

No less than 94% of the annual funding will be dedicated to the components of the HPCS specified in this RFP. Under task orders issued by the Government, the remaining funds will be used by the contractor to refine key areas of the HPCS, or other aspects of GFDL’s computing environment covered under the scope of the contract, that will improve performance, efficiency, and/or usability of the overall system. These areas may include, but will not be limited to, node, disk, or memory upgrades, and visualization, server, and desktop capabilities and the supporting network infrastructure. Key areas will be identified on an annual basis by performance assessments, including an annual system performance review. The Government and the HPCS contractor will work together to identify the necessary hardware and/or software components to be purchased under the contract that will best meet GFDL's computing needs, but the Government will be the sole determiner of which components will be purchased.

Since computing is essential to GFDL's scientific objectives, the HPCS must be characterized by a very high level of reliability and availability. System availability of at least 96% (24 hours/day, 7 days/week) has been the historical goal for GFDL's high-performance computers, and this level of availability must continue to characterize each component of the HPCS and the entire system as a whole.

The final budget for the HPCS procurement has not been finalized, although the targeted total funding level stated in the Project Agreement is approximately $69 million during FY2000-06 and is estimated to be no more than $84 million. GFDL expects that any increases over the targeted funding profile will be used to increase the HPCS computational throughput and other resources needed to provide a balanced system approximately in proportion to the increase in funding.

The initial delivery of the HPCS is expected to be available for contractor testing and preparation for the pre-acceptance Live Test Demonstration (LTD) by the end of summer 2000. The HPCS and the associated support staff will be located in the GFDL buildings on the Forrestal Campus of Princeton University in Plainsboro, NJ.

The basic tenets and provisions of this Statement of Need establish what the Government feels are the minimum acceptable capabilities of the HPCS based on GFDL's experience in performing its mission for a number of years. However, innovation in proposed high performance solutions is encouraged in addressing the needs of the Government. Newer technologies or an approach different from that presented here may provide opportunities to increase performance or enhance efficiency.

C.3 Current Computing Environment

C.3.1 High-performance computing platforms

The key components in GFDL's current computing environment are a Cray T932 shared-memory vector supercomputer with 22 vector processors, a Cray T94 shared-memory vector supercomputer with 4 vector processors, and a scalable, distributed-memory Cray T3E with 128 application processors. A very large data archive is hosted by the T94, the contents of which can also be accessed from the T932, the T3E, and desktop workstations. High-speed interconnections allow very fast data transfers between elements of the central computing facility. The T932, T94, and T3E have similar UNIX operating systems and software development environments, and they all use IEEE 64-bit precision arithmetic. The FORTRAN 90 programming language and parallel code development software is available on all of these systems. The entire computing environment operates 24 hours a day, 365 days a year. Typical operational use time is in excess of 22 hours per day on the T932, 23 hours per day on the T94, and over 23.5 hours per day on the T3E.

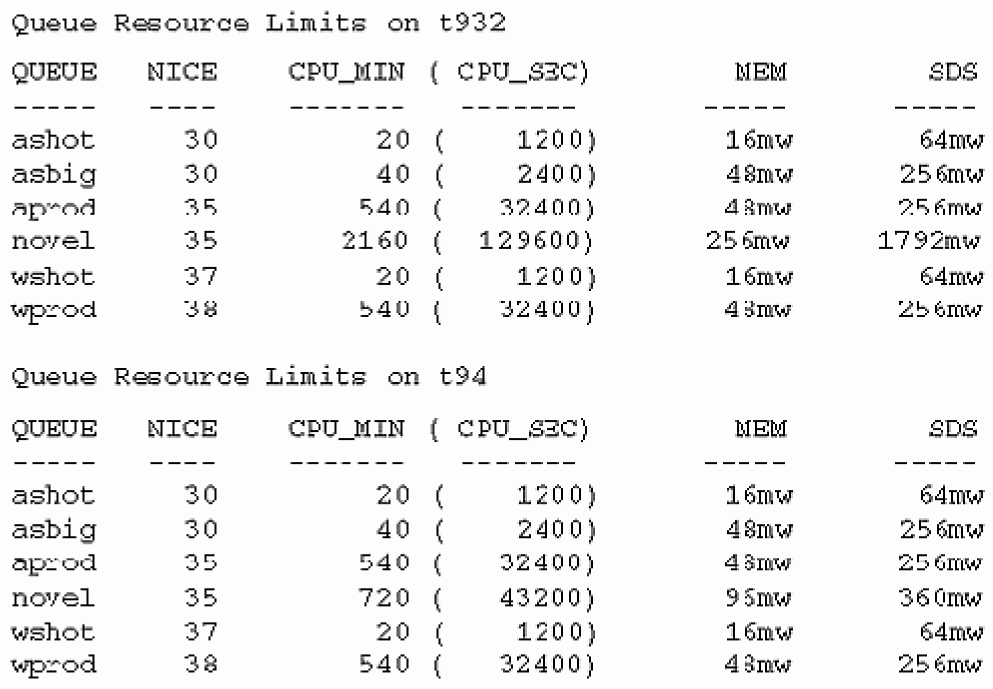
Most of the model runs that require significant computing resources are executed on the T932, which has 4 GB of main memory, 32 GB of secondary data storage (SDS) which is used primarily as secondary memory, and 447 GB of disk storage. Approximately 30 production runs typically execute concurrently on the T932. The production runs consist primarily of a sequence of unitasked jobs (although scientists have also developed several multitasked production applications as well) that resubmit themselves; this sequence may run for months at a time. Typically, each job within a production run will request data from the archive hosted on the T94, transfer it to temporary disk storage on the T932, execute, then transfer output back to the data archive. Memory- and I/O-intensive postprocessing and analysis of this data is usually carried out on the T94, which has 1 GB of main memory, a 4 GB solid state storage device, and 760 GB of disk storage distributed primarily between temporary disk storage and staging disk for data in the tape archive. The T94 also runs approximately 6 concurrent production jobs at night. GFDL has found that dedicating the T932 and T94 resources in this way substantially enhances the efficiency at which these two machines run. Each T90 vector processor sustains an average of about 450 MFLOPS on GFDL's total production workload, so together both T90s provide approximately 12 GFLOPS for these computational tasks.

The T3E has 256 MB of memory on each of its 128 application processors and about 370 GB of disk storage used primarily as scratch space. The T3E’s greatest value is as a development machine in the present computing environment. Users currently employ the T3E to parallelize unitasked codes and port multitasked codes from the shared-memory environment on the T90s to a distributed-memory environment. The T3E's success in this role is evidenced by the 2-4 production jobs that run concurrently on this machine, in a manner similar to that on the T932. The T3E provides about 4 GFLOPS of peak throughput for parallelized production codes.

C.3.2 Batch queuing, scheduling, and accounting

Each group at GFDL is assigned a monthly allocation of T90 CPU hours for batch work. By default, batch jobs are submitted to "allocated" queues. Once a group's monthly allocation has been used, all batch job submitted by that group are forcibly directed to windfall queues. Windfall jobs are run during non-primetime hours (7pm-7am) or when there is not enough allocated work to fully utilize the system.

CPU accounting runs once per day, which limits the granularity for enforcing the monthly allocations. Allocated batch, windfall batch, and interactive CPU time for each group are distinguished using the UNICOS account id feature. The monthly CPU allocation scheme is implemented by a locally-developed wrapper for the NQS qsub command.



```
Queue Resource Limits on t932

QUEUE   NICE    CPU_MIN ( CPU_SEC)        MEM       SDS
-----   ----    -------   -------        -----     -----
ashot     30         20 (    1200)        16mw      64mw
asbig     30         40 (    2400)        48mw     256mw
aprod     35        540 (   32400)        48mw     256mw
novel     35       2160 (  129600)       256mw    1792mw
wshot     37         20 (    1200)        16mw      64mw
wprod     38        540 (   32400)        48mw     256mw

Queue Resource Limits on t94

QUEUE   NICE    CPU_MIN ( CPU_SEC)        MEM       SDS
-----   ----    -------   -------        -----     -----
ashot     30         20 (    1200)        16mw      64mw
asbig     30         40 (    2400)        48mw     256mw
aprod     35        540 (   32400)        48mw     256mw
novel     35        720 (   43200)        96mw     360mw
wshot     37         20 (    1200)        16mw      64mw
wprod     38        540 (   32400)        48mw     256mw
```

Fig. 1 Current batch queues and associated resource limits on the T932 and T94

In Figure 1, queues starting with the letter "a" are allocated queues, and jobs starting with the letter "w" are windfall queues. There is also a "novel" queue that is reserved for jobs with special resource requirements.

To encourage parallelization of GFDL's model codes, monthly CPU allocations are currently not implemented on the T3E.

C.3.3 Disk I/O

The I/O in GFDL's production runs is typically dominated by reading a restart file at the beginning of a run, writing many snapshots of geophysical variables throughout the run (history files), and writing restart data at the end of the run. These runs also write to the standard FORTRAN output units. A restart file is usually a single large binary file, while the history files can be either one large file or many small files (for example, one file written per CPU, typically in netCDF format, on the T3E) that can be merged into a single large netCDF file after the production run finishes. The I/O to large restart files and the efficient creation of many small files on the LSC disk used for temporary storage, as well as the ability of the HSMS (see below) to archive many small files simultaneously, are among GFDL's principal concerns. The following table summarizes current disk performance on all three supercomputers:

Platform	RAID-3 Drives / Channels	JBOD Drives / Channels	Maximum Total Bandwidth
T932	10 / 10	237 / 41	~750 MB/sec
T94	19 / 19 Fibre	6 / 1 Fibre	~900 MB/sec
T3E	6 / 6 Fibre	0 Fibre	~300 MB/sec

C.3.4 Data archive

Centralized data archiving is presently provided by the Cray T94 using the UNICOS Data Migration Facility (DMF). About 500 GB of archive staging disk is hosted by the T94. Two StorageTek (STK) Powderhorn silos contain 8 STK Timberline and 4 STK Redwood tape transports and a library of approximately 6000 IBM 3490E-compatible Timberline tapes (800 MB each), 1100 Redwood tapes (10 GB each), and 2500 Redwood tapes (50 GB each); there are expected to be approximately 1900 free cells in the STK silos at the end of the current contract. Each STK tape transport includes an integrated IBM-compatible ESCON controller. The tape transports provide an aggregate transfer rate of about 88 MB/s between the tape library and the staging disk. Operation of GFDL's computing environment presently requires an average of over 2400 tape mounts per day. During intermittent periods of high activity, up to 350 tape mounts per hour are required, which roughly corresponds to the maximum tape mount rate available with the 8 Timberline tape transports. A data compression factor of about 1.3x with DMF is approximately balanced by a tape fill factor of about 70% to achieve storage at nearly the rated tape capacity.

The number and capacity of files in GFDL's data archive, as a function of file size, are shown in the following figures, in which the blue curve with squares represents cumulative totals of the number of files (left-hand plot) and KB (right-hand plot) in the archive, and the red curves with diamonds represent increments to these quantities for each file size bin.

Note that about 30 TB out of the entire ~100 TB archive is contained in files that are between 200 and 500 MB in size. However, over three-fourths of the number of files in the archive are less than 100MB in size. The following table of daily average tape mount statistics from 1999 shows that about four-fifths of tape mounts are to read and write files less than 80MB in size, indicating that frequent access to files of smaller size is as crucial to GFDL's scientific workload as fast access to files of larger size.

File size	Files Mounted	MB Mounted	Mounts	Files/Mount	MB/Mount
< 80 MB	14898	107146	1280	11.64	83.70
80-160 MB	330	33874	100	3.30	338.74
> 160 MB	510	259080	244	2.09	1061.80

Filesize (KB)	Cumulative #	Δ#	Cumulative KB	Δ KB
4	9639	9639	38556	38556
10	36232	26593	231809	193253
20	49774	13542	433975	202166
50	90841	41067	1798775	1364800
100	126123	35282	4147090	2348315
200	162364	36241	9228524	5081434
500	255587	93223	37170804	27942280
1000	317965	62378	82496383	45325579
2000	459234	141269	296830203	214333820
5000	559702	100468	623949196	327118993
10000	632963	73261	1153257750	529308554
20000	674704	41741	1715057537	561799787
50000	720569	45865	3152423121	1437365584
100000	764415	43846	6592591232	3440168111
200000	866803	102388	21303616561	14711025329
500000	967295	100492	51915683497	30612066936
1000000	998218	30923	74554669526	22638986029
2000000	1009951	11733	88847107619	14292438093
5000000	1012832	2881	96615118130	7768010511
10000000	1013066	234	98103885135	1488767005
20000000	1013117	51	98695633963	591748828
29114976	1013133	16	99083623099	387989136

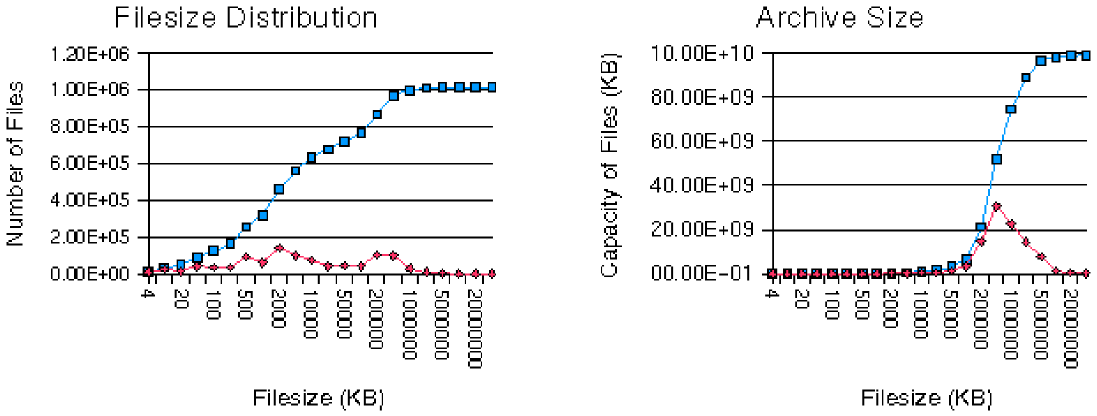


Fig. 2 Number of files and capacity of files as a function of file size distribution

Files less than 80 MB in size are archived to Timberline tapes, files between 80 and 160 MB are archived to 10GB Redwood tapes, and files greater than 160 MB in size are archived to 50GB Redwood tapes; these categories correspond to the rows in the table. Because of the limited number of Timberline tapes available in the current archive, users typically collect many small files (those less than 80MB in size) into a cpio or tar file that is archived on Redwood tapes.

All of GFDL's supercomputers and desktop workstations have access to the entire data archive. The data archive appears as a single directory (/archive) to users on the T90s and workstations. This directory is hosted on the T94. On the T932, read/write access to this directory is provided via NFS, and locally developed scripts employ the rcp command to enhance performance on file transfers. On the T3E, file transfers between /archive and local temporary disk are accomplished with the rcp command. On the workstations, read-only access is provided via NFS.

Presently about 100 TB of data are archived, and the growth rate of the archive has averaged 2.5 TB/month over the last year. The archived data includes restart files, boundary conditions, and initial conditions required to run GFDL's models, history data and analysis files from model runs, datasets for model comparison, and archived source code. Most data is archived in 32-bit IEEE floating-point format in big-endian byte order, which allows the data to be read directly by GFDL's desktop workstations. Restart files are usually stored in 64-bit format in a binary file. Files used for model analysis are written once and read many times, accessed repeatedly over periods ranging from several days to years. Individual users are permitted to stage up to 50 GB of archived files to disk at any given time. The amount of archived data read and written per day is greater than the size of the staging disk space, so in addition to user-directed migration, files are automatically migrated from staging disk to tape when the available space on this disk becomes low.

C.3.5 Desktop workstations

Distributed throughout GFDL are over 110 desktop workstations, mostly Personal Irises, Indigos, and Indigo-2's manufactured by SGI. Most of the workstations are located in scientists' offices; the rest are in public areas. Supporting these workstations are 4 SGI and 2 Sun servers plus a variety of Postscript printers. These workstations are used for data analysis and visualization, as well as day-to-day office tasks such as email and manuscript preparation.

C.3.6 Network configuration

GFDL's current network configuration is discussed in its Computer Users Guide (available by request). Point-to-point HIPPI interconnections at 100 MB/s and 200 MB/s provide fast data transfers between the Cray supercomputers, which are also connected to a computer-room FDDI ring. The facility-wide workstation network is comprised of several 10/100 Mb/s switched Ethernet hubs connected to a Gigabit Ethernet (GBE) switch/router in the computer room. One of the switched Ethernet hubs is connected to the FDDI ring, providing interactive access to the supercomputers from the workstations. Access to the Internet is via a 1.5 Mb/s T-1 communications link to an Internet access provider. There is also a 10 Mb/s connection to the Princeton University network. External information exchange is accomplished using FTP and the World Wide Web; very large data sets are exchanged using 8 mm magnetic tape.

C.3.7 Directory Trees

There are several principal directory trees in use at GFDL. The entire data archive appears as a single /archive directory to the users. Workstation home directories (/home) on a central server can be accessed from any workstation. Separate home directories for the T90s (/t90) are hosted by the T94 and served to the T932 and all workstations. The T3E has its own home directories (/t3e) which are available to the workstations. Temporary directories unique to each interactive session or batch job provide workspace for users on each supercomputer.

All of these computing resources are currently shared by about 100 users at GFDL and nearly 50 other collaborators worldwide. Additional information on GFDL's current computing environment can be found in GFDL's Computer Users' Guide and in GFDL's IT Architecture document (http://www.gfdl.gov/~jps/noaa_hpc/GFDL_IT-Arch.ver1.html).

C.3.8 Government-Furnished Equipment

Government-furnished equipment includes the two STK Powderhorn silos, the 8 STK Timberline and 4 STK Redwood tape transports (including integrated IBM ESCON controllers), and the IBM ESCON director. Two GBE interfaces on GFDL's GBE backbone have been reserved to connect to the HPCS.

C.4 Specifications

C.4.1 Overview

GFDL requires a single contractor to design, install, maintain, and support a high-performance computing system (HPCS) in its laboratory in Plainsboro, New Jersey. The HPCS shall meet the stated objectives and specifications set forth in this Statement of Need and shall include all hardware and software necessary to operate the HPCS as a complete, functional, balanced, and highly reliable system. A single contractor will serve as the point-of-contact for the entire HPCS, even though the HPCS may involve separate subsystems from a number of different vendors.

Fundamentally, GFDL must improve its computing capacity and performance in all aspects of its computing environment in order to fulfill its mission. These aspects include:

- High-performance computing, including large scale computing and analysis capability. Large scale computing is currently supplied by the Cray T932 and T3E, and to some extent by the T94. The HPCS shall include a Large Scale Cluster (LSC) that provides scalable supercomputing capabilities to support GFDL's leading-edge research in geophysical fluid modeling. Analysis capability is currently supplied by the T94 and desktop workstations. The HPCS shall include an Analysis Cluster (AC) that provides efficient execution of I/O-intensive FORTRAN codes and third-party software, which are required to analyze and interpret model output produced by the LSC.
- A Hierarchical Storage Management System, currently supplied by DMF running on the T94 in conjunction with the STK Powderhorn tape libraries and the T94 /archive filesystem. The HSMS shall provide archiving capacity to meet the expected rates of data production on the LSC and AC.
- Shared HPCS resources. A /home directory file server (HFS) shall provide the home directory at login for all computers and workstations at GFDL. High-speed connectivity, currently supplied by the HIPPI connections between the Cray supercomputers and the FDDI, shall provide communication between the LSC, AC, HSMS, and workstations.

- Software for resource management, system administration, and application development. GFDL needs operating systems and/or cluster software that can manage resources on the LSC, AC, HSMS, and /home filesystem. Complete and functional FORTRAN and C application development environments shall also be provided.
- Reliability, availability, and support. The HPCS must continue GFDL's historically high utilization of its computing resources. System reliability, availability, and contractor support are considered fundamental aspects of the HPCS and are an important aspect of the evaluation of any proposed HPCS.

Additionally, the new HPCS shall be able to share power, cooling, and floor space with GFDL's legacy systems while they remain at GFDL. The T932 and T3E are leased through December 2000. Ownership of the T94 will transfer to the government in October 2000.

C.4.2 Requirements

The proposed HPCS shall meet or exceed the following requirements:

C.4.2.1 Large Scale Cluster (LSC)

☐ An LSC of two or more high-performance computers

☐ Completion of the LSC benchmark in no more than 3 hours of wallclock time on the initial LSC

☐ At least one substantial upgrade to the sustained throughput of the LSC during the base contract period

☐ Options to further enhance the LSC throughput after the base contract period

☐ A minimum of 144 GB of total memory for user processes on the initial delivery of the LSC

☐ A minimum of 256 MB per processor

☐ Options for at least 512 MB/processor and 1 GB/processor of memory

☐ A minimum of 3 dTB of formatted disk, exclusive of system disk, residing on a fault-tolerant disk subsystem

☐ A minimum sustained aggregate I/O bandwidth to disk of 4 GB/sec

☐ The ability to store files of up to 100 GB in size on the disk subsystem

☐ The ability to read file formats written by the Analysis Cluster without explicit library calls for data conversion from within the application

☐ Failover capability for job queuing and scheduling and interactive sessions

☐ The ability for the LSC to operate and be repaired in degraded mode

□ The capability to run at least two copies of GFDL's projected largest job when any single computer is unavailable for user jobs

□ Full functionality when the Analysis Cluster is halted and powered off for repair

□ An availability level of 96% on every computer in the LSC

C.4.2.2 Analysis Cluster (AC)

□ An AC of two or more high-performance computers

□ Completion of the AC benchmark in no more than 3600 seconds of wallclock time on the initial AC

□ At least one substantial upgrade to the sustained throughput of the AC during the base contract period

□ One upgrade near the end of September 2001, when the T94 warranty expires

□ Assumption of the T94's maintenance payments or replacement and subsequent disposal of the T94

□ Options to further enhance the AC throughput after the base contract period

□ A logically shared address-space at least 32 GB in size on all computational nodes of the AC

□ A minimum of 64 GB of memory available for user processes on the AC

□ The ability to increase total memory on the AC by increasing the memory on each node or by replacing nodes with ones having larger memory capacities

□ A minimum of 5 dTB of formatted disk, exclusive of system disk, residing on a fault-tolerant disk subsystem

□ A minimum sustained aggregate I/O bandwidth to disk of 6 GB/sec

□ The ability to store files of up to 100 GB in size on the disk subsystem

□ The ability to read file formats written by the LSC without explicit library calls for data conversion from within the application

□ Failover capability for job queuing and scheduling and interactive sessions

□ Full functionality when the LSC is halted and powered off for repair

□ An availability level of 96% on every computer in the AC

C.4.2.3 Hierarchical Storage Management System (HSMS)

□ A 3-tiered storage scheme comprised of disk, nearline storage robotically mounted at high speed, and offline storage

□ Disk required for caching or staging of files residing on a fault-tolerant disk subsystem that is in addition to the required LSC and AC disk

□ A data recovery service for nearline and offline data

☐ Completion of the pre-award archive benchmark in no more than 1800 seconds of wallclock time on the initial HSMS

☐ Government ownership of the HSMS at the end of FY2003

☐ The ability to store at least 1000 dTB of data on nearline media by the end of FY2003

☐ A minimum final total capacity for nearline and offline tiers of 5000 dTB, independent of compression, by the end of FY2006

☐ The ability to store a minimum of 10,000,000 archived files

☐ The ability to store files up to 100 GB in size

☐ A minimum 10 dMB/s sustained single-file transfer rate from nearline storage to disk, independent of compression

☐ A minimum 160 dMB/s aggregate sustained transfer rate between disk and nearline media for access to small frequently-used files, independent of compression

☐ A minimum 200 dMB/s aggregate sustained transfer rate between disk and nearline media for access to large files, independent of compression

☐ A minimum aggregate tape positioning rate of 1600 mounts per hour for access to small frequently-used files

☐ At least one substantial upgrade to the aggregate sustained transfer rate between disk and nearline media

☐ Software that provides automatic migration between data archive tiers based on a combination of access time and file size

☐ Access to the data residing in GFDL's DMF data archive, beginning at the time of the HPCS installation

☐ The ability to read the legacy archive throughout the life of the HPCS

☐ Dedication of legacy tape transports to reading files from the legacy archive unless all legacy data has been offloaded to the new HSMS media

☐ Availability of the /archive filesystem image on the LSC and AC via a protocol such as NFS v.3 or DCE/DFS, or as a shared filesystem

☐ A high-performance file transfer interface such as the UNIX rcp command

☐ Availability of the /archive filesystem image with read/write access on the T932, T94, T3E, and user workstations via NFS v.2 and the UNIX rcp command

☐ Access via NFS v.3 for use by future workstations

☐ Failover capability in the server that manages the data archive

☐ Completion of failover to backup resources within 5 minutes

☐ An availability level of 99.96%

C.4.2.4 Home Directory Filesystem Server (HFS)

☐ A single /home filesystem which will provide the home directory at login for all computers and workstations

☐ Initial delivery of a minimum 1 dTB of formatted disk, exclusive of system disk, residing on a fault-tolerant disk subsystem

☐ At least one substantial upgrade to the disk capacity of the HFS

☐ Transfer of all data residing in GFDL's workstation home directories at the time of the HPCS installation

☐ Transfer of all data residing in the T90 and T3E home directories at the time the T932 and T3E are de-installed

☐ Availability of the /home filesystem on the LSC and AC via a protocol such as NFS v.3 or DCE/DFS, or as a shared filesystem

☐ Availability of the /home filesystem with read/write access on the T932, T94, and T3E, and user workstations via NFS v.2 and the standard UNIX rcp command

☐ Implementation of per-user and per-group disk space quotas for the /home filesystem. The quota and current use shall be viewable via user commands on the LSC, AC, the T932, T94, and T3E, and user workstations.

☐ Failover capability in the server that manages the /home directories

☐ Completion of failover to backup resources within 5 minutes

☐ An availability level of 99.99%

C.4.2.5 Connectivity

☐ The connection of the LSC, AC, HSMS, and HFS to GFDL's Gigabit Ethernet (GBE) workstation backbone at a minimum of GBE speeds

☐ High-performance transfer of files in the HPCS data archive to GFDL's servers and workstations via the two government-furnished GBE interfaces

☐ Continued access to the HPCS when one of the two government-furnished GBE interfaces fail

☐ High-performance file transfers at GBE speeds or better between the LSC, AC, HSMS, and HFS

☐ An upgrade to GFDL’s access to the Internet to a minimum of T-3 or it’s equivalent

☐ Availability of the /archive and /home filesystems on the LSC, AC, and GFDL workstations

C.4.2.6 Software

☐ HPCS software that meets all government standards

☐ System software on the LSC and AC listed as required in C.4.6.1

☐ Resource management software on the LSC, AC, HSMS, and HFS listed as required in C.4.6.2

☐ Applications software for the programming environment on the LSC and AC listed as required in C.4.6.3

☐ X-windows applications software for the AC listed as required in C.4.6.4

C.4.2.7 Reliability, Availability, and Support

☐ An availability level of 96% for the entire HPCS

☐ A designated point of contact to request maintenance

☐ An escalation procedure that allows the Government continuous telephone coverage should the designated point of contact be unavailable

☐ Preventative maintenance that is completed before the start of GFDL primetime

☐ Benchmark modifications made by mutual agreement

☐ An uninterruptable power supply (UPS) for all components of the HPCS

☐ A minimum of 2 software engineers on and a minimum of 1 hardware engineer on site, with at least one software engineer and one hardware engineer available during GFDL primetime, five days per week

☐ Additional on-call support 24 hours per day, 7 days per week, with a 2-hour response time

☐ An itemized list of all contractor-supplied hardware and software items, and documentation of these items in printable electronic form

☐ Training for approximately 30 GFDL computer specialists and operators

☐ Training for approximately 100 applications programmers

☐ Pre-delivery access to systems similar to those proposed for the HPCS

C.4.2.8 Facilities

☐ An HPCS that meets the requirements for power, cooling, and floor space discussed in C.4.7.2.2 through C.4.7.2.4

☐ Design and construction of rooms adjacent to the Computer Room as outlined in C.4.7.5

C.4.3 Desired features

The following features are considered desirable on the proposed HPCS:

C.4.3.1 Large Scale Cluster (LSC)

☐ Binary compatibility of all processors

☐ The identical configuration of all computational nodes

☐ The identical OS level on all processors and computational nodes

☐ The ability for a single message-passing application to access at least one half of the computational nodes

☐ Linear scaling of memory with throughput on systems that exceed the minimum throughput requirements

☐ Interactive resources that are isolated from the batch production resources

☐ The ability to reassign interactive resources to the batch production jobs during non-primetime hours without a reboot of the entire LSC

☐ The ability to test OS and application software upgrades in isolation from the interactive and production resources

☐ User login to a single hostname

☐ Failover to binary-compatible processors running the identical OS level

☐ Automatic rerun of batch jobs upon resource failure

☐ Loss of only those interactive jobs hosted on the failed resources

☐ All batch jobs remaining visible to job status commands until complete, even if the computer running the job has crashed

☐ No single point of failure

C.4.3.1 Analysis Cluster (AC)

☐ Binary compatibility of all processors

☐ The identical configuration of all computational nodes

☐ The identical OS level on all processors and computational nodes

☐ The ability for a single message-passing application to access at least one half of the computational nodes

☐ User login to a single hostname

☐ The ability to test OS and application software upgrades in isolation from the interactive and batch resources

☐ Failover to binary-compatible processors running the identical OS level

□ Automatic rerun of batch jobs upon resource failure

□ Loss of only those interactive jobs hosted on the failed resources

□ All batch jobs remaining visible to job status commands until complete, even if the computer running the job has crashed

□ No single point of failure

C.4.3.3 Hierarchical Storage Management System (HSMS)

□ Balance between robot performance and tape positioning rate

□ A method for determining the location of users' files within the storage hierarchy

□ User-specified migration between tiers through a single software interface

□ An HSMS that can send files from tape directly to different destinations over the network

□ The ability for users to group related files and directories on a single tape volume

□ Accounting for the HSMS that reports individual usage at the group and user level of all storage tiers

□ Presentation of the new and legacy data archive to the user as one /archive filesystem image

C.4.3.4 Home Directory File Server (HFS)

□ Access via NFS v.3

C.4.3.5 Connectivity

C.4.3.6 Software

□ System software on the LSC and AC listed as desired in C.4.6.1

□ Resource management software on the LSC, AC, HSMS, and HFS listed as desired in C.4.6.2

□ Applications software for the programming environment on the LSC and AC listed as desired in C.4.6.3

□ X-windows applications software for the AC listed as desired in C.4.6.4

The requirements and desirable features on the HPCS are described in more detail below.

C.4.2 High-Performance Computing

The HPCS shall provide high-performance computing resources for large-scale computing and analysis capabilities.

C.4.2.1 Large Scale Cluster (LSC)

Scalable supercomputing capabilities shall be provided by a Large Scale Cluster (LSC) of two or more high-performance computers, as defined in C.6. GFDL desires binary compatibility of all processors and desires the identical configuration of all computational nodes within the LSC. Node homogeneity will be evaluated. The identical OS level is also desired on all processors and computational nodes. It is desirable that a single message-passing application be able to access at least one half of the computational nodes on the LSC.

C.4.2.1.1 LSC performance

The LSC shall provide a substantial increase in sustained throughput over that provided by GFDL's current Cray supercomputers described in C.3.1. Sustained throughput will be measured by an LSC throughput benchmark (J.#.#) comprised of concurrently-run parallelized codes sampled from GFDL's expected future workload. This benchmark shall run in no more than 3 hours of wallclock time on the initial delivery of the LSC. In addition, the scalability of the LSC will be measured by a benchmark designed to reveal the performance and scaling characteristics of individual codes as they are executed on different processor counts. Kernels and microbenchmarks may also be used for evaluating important system operations and interpreting benchmark performance.

At least one substantial upgrade to the sustained throughput of the LSC, as measured by the throughput benchmark, shall be offered during the base contract period. Additional upgrades may be proposed, but the evaluation will favor a limited number of substantial upgrades. Options to further enhance the LSC throughput after the base contract period shall be offered, as discussed in section C.5.

C.4.2.1.2 LSC memory

A minimum of 144 GB of total memory shall be available for user processes on the initial delivery of the LSC. These numbers represent a linear scaling with throughput of the combined (36 GB) main memory and solid-state storage capacity present on the T932. It is desirable that the memory scale linearly with throughput on systems that exceed the minimum throughput requirements. A minimum of 256 MB/processor is required, and mandatory options for at least 512 MB/processor and 1 GB/processor of memory shall be offered.

C.4.2.1.3 LSC disk I/O

The initial delivery of the LSC shall have fast access to a minimum of 3 dTB of formatted disk, exclusive of system disk (which may include, for example, RAID parity disks, swap partitions, and dump partitions). This disk space shall reside on a fault-tolerant disk subsystem. Performance in degraded mode (e.g., when one or more spindles in a RAID-configured disk subsystem ceases to function) will be evaluated.

Most of this disk space will be used as temporary storage space for files accessed by the production workload. GFDL requires a minimum sustained total I/O bandwidth of 4GB/sec to support the suite of jobs in the production workload. The number of channels to this disk will be evaluated. The minimum capacity and sustained total I/O bandwidth are based primarily on an extrapolation from the rates at which GFDL currently writes data to its temporary storage space to the expected rates at which data will be written in GFDL's projected production workload. The effectiveness with which the I/O subsystem on the LSC can access files, particularly the restart and history files described in section C.3.1, will be evaluated. The disk subsystem shall be able to store files of up to 100 GB in size.

C.4.2.1.4 Interactive and testing resources on the LSC

The Government desires resources for interactive work that are isolated from the batch production resources. It has been GFDL's experience that the nature of interactive work creates resource contention with production jobs, and production jobs slow interactive response time. Credit will be given for the ability to reassign interactive resources to the batch production jobs during non-primetime hours without a reboot of the entire LSC. The interactive resources will be evaluated by performing compilations, debugging, and file management tasks during scripted interactive sessions on the LSC during the pre-award LTD, concurrent with the execution of the throughput benchmark. The Government desires the ability to test OS and application software upgrades in isolation from the interactive and production resources on the LSC. User login to a single hostname is desirable.

C.4.2.1.5 LSC reliability, availability, and support

Failover capability for job queuing and scheduling and interactive sessions shall be provided, as shall the capability for the LSC to operate and be repaired in degraded mode. It is desirable that failover be to binary-compatible processors running the identical OS level. It is desirable that when any set of resources in the LSC fails (such as disk or memory), batch jobs using those resources are rerun automatically, only interactive sessions hosted on the failed resources are lost, users continue to be able to login interactively, and all batch jobs remain visible to job status commands until complete, even if the computer running the job has crashed. When any single computer is unavailable for user jobs, the remaining resources shall be capable of running at least two copies of our projected largest job (FMS N270L40, which uses an estimated 15 GB of memory). The LSC shall be fully functional when the Analysis Cluster is halted and powered off for repair. It is desirable that the LSC have no single point of failure. The Government requires an availability level of 96% on every computer in the LSC.

C.4.2.2 Analysis Cluster (AC)

Enhanced analysis capability shall be provided by an Analysis Cluster (AC) of two or more high-performance computers that efficiently execute I/O-intensive FORTRAN codes and third-party software. Typically, FORTRAN codes are created and modified frequently, and their potentially short lifetime makes them, as well as any third-party software, unlikely candidates for parallelization by GFDL scientists. It is desirable that the AC feature binary compatibility of all processors, the identical configuration of all computational nodes, the identical OS level on all processors and computational nodes, and the ability for a single message-passing application to access at least one half the computational nodes.

C.4.2.2.1 AC performance

It is desirable that the sustained throughput available on the AC maintain a balance with that on the LSC. Sustained throughput will be measured by a throughput benchmark comprised primarily of unitasked codes characterized by a higher I/O:compute ratio than in the production codes of the LSC throughput benchmark. One example would be the combination of many small history files into one large netCDF file, as described in section C.3.3. The benchmark shall run within # hour of wallclock time. The wallclock performance of single jobs may also be used to assess performance. Kernels and microbenchmarks may be used for evaluating important system operations and interpreting benchmark performance.

An aggressive upgrade path, determined through discussions between contractors and the government, shall be provided for the AC. At least one substantial upgrade to the sustained throughput of the AC, as measured by the throughput benchmark, shall be offered during the base contract period. Additional upgrades may be proposed, but the evaluation will favor a limited number of substantial upgrades. One upgrade shall be scheduled near the time that the warranty on the T94 expires at the end of September 2001. The contractor shall either assume the T94's maintenance payments or replace the T94 with an upgrade to the AC. The upgrade itself shall provide more sustained throughput than the T94, as measured by the analysis throughput benchmark. The contractor will be responsible for disposing of the T94. Options to further enhance the AC throughput after the base contract period shall be offered, as discussed in section C.5.

C.4.2.2.2 AC memory

GFDL requires a logically shared address-space, at least 32 GB in size, preferably in hardware, on all computational nodes of the AC. A minimum of 64 GB of memory shall be available for user processes on the AC. GFDL requires the option to increase total memory on the AC by increasing the memory on each node (rather than just adding more nodes) or by replacing nodes with ones having larger memory capacities.

C.4.2.2.3 AC disk I/O

The AC shall have fast access to at least 5 dTB of formatted disk, exclusive of disk used for system use (which may include, for example, RAID parity disks, swap partitions, and dump partitions). This disk space shall reside on a fault-tolerant disk subsystem. Performance in degraded mode will be evaluated. Most of this disk space will be used as temporary storage space for files accessed by the analysis workload. GFDL requires a minimum sustained aggregate I/O bandwidth to disk of 6 GB/sec to support the suite of jobs in the analysis workload. The minimum capacity and sustained total I/O bandwidth is based primarily on an estimate of the size of model datasets produced by GFDL's projected production workload. The disk subsystem shall be able to store file of up to 100 GB in size.

Applications on the AC shall be able to read file formats written by the LSC, and vice versa, without explicit library calls for data conversion from within the application. Formats frequently used in LSC applications include FORTRAN sequential and direct-access files and C text and binary stream files.

C.4.2.2.4 AC reliability, availability, and support

User logins to a single hostname for the AC is desirable. The ability to test OS and application software upgrades in isolation from the interactive and batch resources on the AC is also desirable.

The AC shall provide failover capabilities functionally equivalent to those provided for the LSC. Failover to binary-compatible processors running the identical OS level is desirable. Failover capability for job queuing and scheduling and interactive sessions shall be provided. The Government desires that when any set of nodes in the AC fails, batch jobs using those nodes are rerun automatically, only interactive sessions hosted on the failed nodes are lost, and users should still be able to login interactively. It is desirable that all batch jobs remain visible to job status commands until complete, even if the computer running the job has crashed. The AC shall be fully functional when the LSC is halted and powered off for repair. It is desirable that the AC have no single point of failure. GFDL requires an availability level of 96% on every computer in the AC.

C.4.3 Hierarchical Storage Management System

GFDL requires a 3-tiered storage scheme for its data archive, comprised of disk, nearline storage (robotically mounted at high speed), and offline storage (with an emphasis on high reliability) that can effectively satisfy the requests for scientific data that permeates GFDL's scientific workload, as discussed in section C.3.4. If disk is required for caching or staging of files within the HSMS, it shall be fault-tolerant and in addition to the required LSC and AC disk specified in section C.4.2. The offline data may be mounted either robotically or manually. For nearline and offline data, a data recovery service shall be provided by the contractor in the event of media failure. Both tape reliability and the data recovery service will be evaluated.

C.4.3.1 HSMS performance

HSMS perfomance will be evaluated by an archive benchmark that will transfer a mix of large and small files across the disk and nearline tiers. The complete benchmark shall run on the proposed HSMS in no more than 1800 seconds of wallclock time. Kernels and microbenchmarks that test additional aspects of the HSMS may also be used for evaluation.

At HSMS installation, a minimum performance for reading files from the legacy archive shall also be demonstrated. Data movement between nearline and offline tiers will be evaluated when the offline tier is delivered.

At least one substantial upgrade in the performance of the HSMS is required during the base contract period.

The HSMS becomes the property of the government at the end of FY2003.

C.4.3.2 HSMS capacity

A schedule for delivery of nearline tape capacity and the offline tier will be mutually agreed upon by the contractor and the Government. No offline storage is required initially. By the end of FY2003, the nearline tier shall be able to store at least 1000 dTB of data, independent of compression. A minimum final total capacity for nearline and offline tiers of 5000 dTB, independent of compression, is required by the end of FY2006. The data archive shall be able to store a minimum of 10,000,000 archived files. The HSMS shall be able to store files of up to 100 GB in size.

C.4.3.3 File migration on the HSMS

GFDL requires a minimum 10 dMB/s sustained single-file transfer rate from nearline storage to disk, independent of compression. This is approximately the current transfer rate between the Timberlines/Redwoods and the archive staging disk on the T94. The bandwidth between disk and the nearline tier should accommodate GFDL's typical migration patterns as discussed in section C.3.1. A minimum 160 dMB/s aggregate sustained transfer rate between disk and nearline media is required for access to small frequently-used files (as discussed in section C.3.1), independent of compression. Further, a minimum aggregate tape positioning rate, defined in Section C.6, of 1600 mounts per hour shall be provided for access to small frequently-used files. This is approximately twice that currently available on the current Timberline tape transports. Robot performance in mounting and unmounting tapes shall balance the tape positioning rate, and will be evaluated using the archive benchmark results and the proposed tape positioning rate. In addition to the 160 dMB/s aggregate sustained transfer rate for small frequently-used files, a minimum 200 dMB/s aggregate sustained transfer rate between disk and nearline media is required for access to large files (as discussed in section C.3.1), also independent of compression.

C.4.5.4 HSMS software

The HSMS software shall provide automatic migration between data archive tiers based on a combination of access time and file size. A method for determining the location of users' files within the storage hierarchy is desirable. User-specified migration between tiers through a single software interface is also desirable. Credit will be given for an HSMS that can send files from tape directly to different destinations over the network rather than to just one archive staging filesystem. Credit will also be given if users can group related files and directories on a single tape volume. Accounting for the HSMS that reports individual usage at the group and user level of all storage tiers is desirable.

C.4.5.5 Legacy archive

All of the data residing in GFDL's DMF data archive, currently hosted on the T94, shall be readable by users throughout the base contract period. It is desirable that the new and legacy data archive be presented to the user as one /archive filesystem image. Possible solutions include maintaining the T94 throughout the life of the HPCS, maintaining a different DMF archive server throughout the life of the HPCS, or moving the legacy data onto the new HSMS media.

The T94 is government-owned and includes hardware maintenance through September 2001. If the T94 is used to serve the legacy archive, the contractor shall provide system administration for the T94, including DMF, after the current contract expires at the end of October 2000. If the T94 is still in use to serve the legacy archive at the expiration of it’s warranty at the end of September 2001, the contractor shall assume the hardware maintenance payments or provide functionally equivalent access to the legacy archive.

The two STK Powderhorn tape libraries and the 8 Timberline and 4 Redwood tape transports, including integrated IBM ESCON controllers and the IBM ESCON director, currently owned by GFDL will be provided as government-furnished equipment. There are four ESCON interfaces available on the ESCON director. After September 2000, the contractor will be responsible for maintaining these tape libraries, drives, and the ESCON director while in use. The Timberline and Redwood tape transports shall be dedicated to reading files from the legacy archive unless all legacy data has been offloaded to different media. The legacy media and tape drive bandwidth do not count toward the required archive capacity and inter-tier bandwidth.

If the legacy archive is no longer served by the T94 before the T932 and T3E are de-installed, the T932 and T3E shall access the new HSMS archive via point-to-point HIPPI connections.

The LSC and AC shall be capable of accessing the data in the legacy archive. If the T94 is used to serve the legacy archive, the LSC and AC shall connect to it via a high-performance HIPPI or FDDI interface to the T94.

C.4.5.6 HSMS reliability, availability, and support

The /archive filesystem image shall be available on the LSC and AC via a protocol such as NFS v.3 or DCE/DFS, or as a shared filesystem. A high-performance file transfer interface, such as the UNIX rcp command, is also required on the LSC and AC. The /archive filesystem image shall also be available with read/write access on the T932, T94, T3E, and user workstations via NFS v.2 and the UNIX rcp command. This allows locally developed file transfer scripts to execute correctly. NFS v.3 is also desired for use by future workstations.

GFDL requires failover capability in the server that manages the data archive and requires an availability level of 99.96% for the data archive. Failover to backup resources shall be complete within 5 minutes.

C.4.4 Home Directory Filesystem Server (HFS)

GFDL requires a single high-availability /home filesystem which will provide the home directory at login for all computers and workstations at GFDL. This disk space shall reside on a fault-tolerant disk subsystem whose performance in degraded mode will be evaluated. A minimum of 1 dTB user-accessible formatted disk shall be delivered initially. At least one substantial upgrade to the disk capacity of the HFS is required during the base contract period.

The performance of the /home filesystem server (HFS) will be evaluated at the pre-award LTD by a benchmark that transfers files between the HFS and the LSC and AC.

All of the data residing in GFDL's workstation home directories at the time of the HPCS installation shall be transferred to the new /home filesystem. All of the data residing in the T90 and T3E home directories shall be transferred to the new /home filesystem when the T932 and T3E are de-installed.

The /home filesystem shall be available on the LSC and AC via a protocol such as NFS v.3 or DCE/DFS, or as a shared filesystem. The /home filesystem shall also be available with read/write access on the T932, T94, and T3E, and user workstations via NFS v.2 and the standard UNIX rcp command. NFS v.3 is required for use by future workstations.

The /home filesystem server shall implement per-user and per-group disk space quotas for the /home filesystem. The quota and current use shall be viewable via user commands on the LSC, AC, the T932, T94, and T3E, and user workstations.

GFDL requires failover capability in the HFS and requires an availability level of 99.99% for the /home filesystem. Failover to backup resources shall be completed within 1 minute on the HFS.

C.4.7 Connectivity

GFDL requires that the LSC, AC, HSMS, and HFS connect to GFDL's Gigabit Ethernet (GBE) workstation backbone at a minimum of GBE speeds. The two government-furnished GBE interfaces shall provide high-performance transfer of files in the HPCS data archive to GFDL's servers and workstations and continued access to the HPCS when one of the two interfaces fails. High-performance file transfers at GBE speeds or better shall be provided between the LSC, AC, HSMS, and HFS. An upgrade to GFDL's access to the Internet, to a minimum of T-3 or it's equivalent, shall be provided.

The /archive and /home filesystems shall be available within the HPCS as discussed in sections C.4.5 and C.4.6.

C.4.8 Software

The HPCS software shall meet all government standards.

C.4.8.1 Operating system software

System software required for the LSC and AC includes

- UNIX-like or licensed UNIX OS
- X11 windowing
- NFS v.2, NIS, DNS
- ftp, rcp, Telnet, BSD lpd
- performance monitoring

Credit will be given for NFS v.3 on the LSC and AC.

System software desired for the LSC and AC includes

- The same or functionally equivalent OS on the LSC and AS.
- Operator-directed checkpoint/restart capability

GFDL has used operator-directed checkpoint/restart as a critical tool in managing its supercomputing systems over the last 10 years. The level of checkpointing available on the LSC and AC will be evaluated. Credit will be given for operator-directed checkpointing without user intervention.

C.4.8.2 Resource management software

The efficient operation of the HPCS requires resource management that will facilitate the use of the LSC and AC by GFDL's scientists as well as providing maximum throughput for their workload. The ease with which the resource management software allows users to manage their workload will be evaluated. GFDL may wish to implement a variety of charge-back algorithms for monthly processor time or enforce different resource allocations, including disk and tape quotas, at the group and user level on each of these HPCS subsystems. Credit will be given for the ability to set job resource limits for the number of processors, CPU time, and memory per process, depending on the project or job class.

GFDL requires resource management software that provides:

- Failover capabilities
- Batch queuing and scheduling
- Job accounting on the LSC and AC that can set resource limits for processes and jobs
- Accounting software for the LSC, AC, HSMS, and HFS that reports resource usage at the group and user level
- System activity monitoring software on the LSC and AC that shows user and system CPU utilization and I/O wait time
- Automated network backup for the system disks on the LSC, AC, HSMS, and the entire /home filesystem on the HFS

GFDL desires resource management software on the LSC, AC, HSMS, and HFS that provides:

- Resubmission of batch jobs upon resource failure
- Enforcement of resource limits on jobs, processes, and user projects
- Management of the interactive and batch resources and the resources devoted to OS and software upgrades
- The status of jobs over the entire LSC or AC
- A single hostname for the LSC or AC to the user community.
- The ability to set job resource limits for the number of processors, CPU time, and memory per process, depending on the project or job class
- Batch queuing and scheduling that provides
    - The capability of batch queuing and scheduling to be based on total disk, memory, processors, and processor time usage
    - A user interface, an operator interface, and load-balancing capability, on the LSC and AC
    - Spooling of job scripts and printed output
    - The ability to request a range of processors on which to run a job
    - The ability to schedule jobs within a node
- Job accounting on the LSC and AC that provides total and high-water-mark resource usage (including nodes, memory, and disk)
- Accounting software that can create separate projects within a group and report the resource usage for each project

- System activity monitoring software on the LSC and AC that can produce a unified report or display user and system CPU utilization and I/O wait time of for the LSC or AC as a whole

C.4.8.3 Programming environment software

Required applications software for the LSC and AC programming environments includes:

- FORTRAN 90/95, C, C++ programming environments, including
  - macro preprocessing
  - source-level debuggers
  - interactive performance profilers
  - support for 64-bit integers
  - support for reading and writing big-endian and little-endian data
  - support for reading and writing 32-and 64-bit IEEE floating-point formats in I/O operations
  - facilities for source code management, including the "make" utility.

  Credit will be given for a single application programming environment common to both the LSC and AC.
- netCDF and UDUNITS libraries (available at http://www.unidata.ucar.edu/)
- NAG numerical libraries
- hardware performance monitoring for all HPCS systems
- MPI, MPI-2
- data conversion libraries (including endian, IEEE, and proprietary data format conversions)

Desired applications software for the LSC and AC programming environments includes:

- parallelized, optimized numerical libraries on the LSC and AC
- optimized (and possibly proprietary) I/O libraries.

C.4.8.4 X-windows applications software

Required X-windows applications software for the AC includes:

- Matlab (http://www.mathworks.com)
- Ferret (http://ferret.wrc.noaa.gov/Ferret)
- Mathematica (http://www.mathematica.com)
- IDL (http://www.rsinc.com/idl/index.cfm)
- GrADS (http://grads.iges.org/grads)

- S-Plus (http://www.mathsoft.com)
- NCAR graphics (http://ngwww.ucar.edu)

Desirable X-windows applications software for the AC includes:

- NAG Iris Explorer

C.4.9 Reliability, Availability, and Support

The HPCS shall continue GFDL's historically high utilization of its computing resources. System reliability, availability, and contractor support are considered fundamental aspects of the HPCS and are an important aspect of the evaluation of any proposed HPCS.

C.4.9.1 Downtime

Downtime is that period of time when all of an HPCS component's scheduled workload cannot be accomplished due to a malfunction in the Contractor-maintained HPCS hardware or software, or when the HPCS or a component of the HPCS is released to the Contractor for Remedial Maintenance (RM).

Null time is that period of time when the scheduled workload cannot be accomplished due to environmental failure, such as loss of electric power or cooling, or recovery from environmental failure. Null time will not be counted as downtime.

The Government shall be the sole determiner of whether any HPCS component's scheduled workload can be accomplished. At the discretion of the Government, downtime is accumulated on the entire HPCS when the LSC, AC, HSMS or HFS is down. GFDL requires at least 96% availability of the entire HPCS each month.

The Contractor shall provide the Government with a designated point of contact to request maintenance. The Contractor shall maintain escalation procedures that allow the Government continuous telephone coverage should the designated point of contact be unavailable.

A component's downtime shall commence at the time the Government makes a bona fide attempt to contact the Contractor at the designated point of contact. At this time the Government will begin a log of the problem which will be completed and signed by both the Government and the Contractor when the problem is resolved. Information to be entered into the log will be determined by the Government.

A component's downtime shall exclude any time in which the Government denies the Contractor maintenance personnel access to the malfunctioning hardware and/or software.

A component's downtime shall end when the computer is returned to the Government in operable condition as determined by the Government, ready to perform all of the scheduled workload.

Preventative maintenance is to be completed before the start of GFDL primetime (7am-7pm), and will count as downtime.

Preparation for post-upgrade LTDs, including any benchmark development, will count as downtime.

C.4.9.2 Availability

Proposed throughput benchmark performance levels will be combined with the proposed availability level to determine a measure of overall proposed system-life throughput for the LSC and for the AC. The actual throughput will be measured on a periodic basis, to be mutually determined by the government and contractor, by combining the demonstrated benchmark performance with the operational use time on the LSC and on the AC.

Shortfalls in throughput on the LSC or on the AC would have to be made up with new equipment brought in at no additional cost. Using the demonstrated benchmark performance on the upgraded system, the government will calculate how long the upgrade shall stay in place to compensate for the shortfall in throughput. This will be rounded up to a multiple of 6-month intervals to minimize system disruption.

Accumulated computational cycles (in CPU-hours) that are lost when jobs are lost due to system failure or system reboot will not count toward the system throughput calculation done for the mutually determined period. If the accounting software cannot report the accumulated computational cycles, it will be assumed that 4 CPU-hours were lost for each processor on which the job ran.

An uninterruptable power supply (UPS) is required for all components of the HPCS. These will provide sufficient power during environmental failure to gracefully shut down the HPCS.

C.4.9.3 Support

GFDL requires a minimum of 2 software engineers on site (to provide a comprehensive system administration service) and a minimum of 2 hardware engineers on site, with at least one software engineer and one hardware engineer available during GFDL primetime, five days per week. Additional on-call support shall be provided 24 hours per day, 7 days per week, with a 2-hour response time. The Government reserves the right to substitute either or both hardware engineers with software engineers during the life of the contract on an as-needed basis. Problem escalation procedures will be evaluated.

GFDL needs an itemized list of all contractor-supplied hardware and software items, and documentation of these items, in printable electronic form.

Training shall be provided for approximately 20 GFDL computer specialists and operators in the following areas on the LSC, the AC, HSMS, and the HFS :

- system administration and tuning
- hardware operation and system overview

Training shall be provided for approximately 100 applications programmers in the following areas on the LSC and the AC:

- application and shell programming
- programming languages and tools
- HSMS software

Contractors will provide GFDL with a list of additional potential training topics.

C.4.9.4 LTD

GFDL requires a pre-award live test demonstration (LTD) on the HPCS hardware offered initially. Instructions for the LTD are provided in Section J. Credit will be given for an LTD performed on hardware identical to that offered for initial delivery. GFDL requires an LTD on the hardware offered for all upgrades at the time of each upgrade. Within one month after award of contract, the government needs pre-delivery access to systems similar to those proposed for the HPCS to develop and test codes and scripts. Delivery shall be complete within 60 days of award.

C.4.10 Facilities Description and Requirements

C.4.10.1 Overview

The HPCS will be installed in the Computer Room of the GFDL Computer Building, which is located behind the main GFDL building. These two buildings, together with ten acres of land, make up the GFDL Complex, which is located on the B Site of Princeton University=s Forrestal Campus in Plainsboro Township, Middlesex County, New Jersey. The Complex is leased from Princeton University through a triple-net lease agreement under which the University owns the buildings and land, but the Government has primary responsibility to maintain the buildings and the equipment therein. Princeton University provides water and sewage utilities and maintains campus roads and grounds; however, GFDL purchases electricity and natural gas directly from the local utility, Public Service Gas and Electric. The Government obtains day-to-day facility support from the University on a pay-as-you-go basis.

The GFDL Computer Building was constructed in 1980 in order to provide a dedicated facility to house the Laboratory's central computer system and associated equipment and to provide office space for the GFDL Computer Systems and Operations staff and contractor support personnel. The new building was designed to provide sufficient raised floor space in the computer room to allow GFDL to operate two generations of systems concurrently, together with associated data archival storage, during transitions from one generation system to the next. When the Laboratory's UNIX workstation network was installed throughout the Laboratory buildings in the late 1980's, GFDL installed network servers and routers in the Computer Building as well.

The Initial System will be installed and operated in parallel with the full SGI/Cray configuration. The Government's intent is to continue the lease and maintenance of the currently configured T932 and T3E systems and associated equipment until thirty days after the Government=s acceptance of the new system or December 31, 2000, whichever date is later. SGI/Cray Research will remove this equipment from the premises within thirty days after the end of the lease. As indicated in C.5.2.2.1 of the Statement of Need, the T94 system and associated equipment will continue to operate through September 30, 2001. SGI/Cray will provide maintenance of this equipment under an extended warranty through this date.

If the winning contractor is unable to successfully complete acceptance of the Initial System thirty days prior to December 31, 2000, the Government will reduce the funds available within the contract in that year in order to pay the lease, maintenance, and support of the T932/T3E equipment. The cost per month has been negotiated with SGI/Cray to be $304,000. However, SGI/Cray has informed the Government that it will only support these systems through March 31, 2001. If the Government does not accept the Initial System thirty days prior to this date, the contractor shall provide the GFDL users with access to computational and archival storage resources that are at least equivalent to the computing environment of the T932 and T3E systems. This should be done in a manner that will allow GFDL users to continue their computational research without interruption or degradation, beginning April 1, 2001.

C.4.10.2 Available Power

The existing electrical service to the GFDL Complex and the rest of Princeton University's B-Site of Forrestal Campus is served from a PSE&G utility substation, a 2000 kVA 13200 to 4160 volt oil filled transformer and 4160 volt switchgear*.* This substation is located on the southwest corner of the Computer Building directly adjacent to the parking lot. This substation is assumed to have been sized based on the power requirements, including both the GFDL Complex and the rest of the B-Site of Forrestal Campus. However, the contractor will need to install an additional transformer and associated equipment if the available power to the Computer Building becomes inadequate for the new equipment to be installed in the Computer Room. In this case, the contractor shall ask the owner, Princeton University, to request that PSE&G upgrade the electrical service as required. An underground 4160-volt feeder, dedicated to the GFDL Complex, is routed from the utility substation to separate building substations located within the GFDL Main Building and Computer Building. The Main Building substation (1500 kVA 4160 to 480 volt transformer) provides power to the Main Building and to the primary chilled water plant for the GFDL Complex. The subsequent analysis will only deal with the Computer Building substation, which is to be the only source of power for the new equipment to be installed in the Computer Room. The Computer Building substation is located in the Transformer Room, the location of which is shown in Figure 1. This substation is comprised of a 4160-volt air interrupter switch, 1500-kVA silicone-filled transformer, and a 2000-ampere, 480/277-volt main switchboard. This equipment was installed around 1980 when the Computer Building was constructed. The substation provides power to the Computer Building and to a 225-ton chiller and cooling tower (referred to as Chiller #4) that will serve primarily as a backup to the primary chiller plant. The lighting, large mechanical equipment (backup chiller, pumps, A/C units, etc) and some computer equipment are served at 480/277 volts via panel boards located throughout the building. The building receptacles, small motors, desktop computers, computer room equipment, and similar loads are served at 208/120 volts via step-down transformers and panel boards. The new Initial System, which is to be operated in parallel with the existing SGI/Cray equipment, may require additional 480/277- and 208/120-volt panel boards to support the new equipment as part of site preparation.

A demand meter is installed as part of the Computer Building substation. This meter was set to record the building demand for a period of one week. The meter indicated a maximum building demand of 595 kVA. This load was taken at a time of the year when the building's cooling system was not in operation. The demand load associated with this cooling system is approximately 230 kVA. Therefore, the worst case building

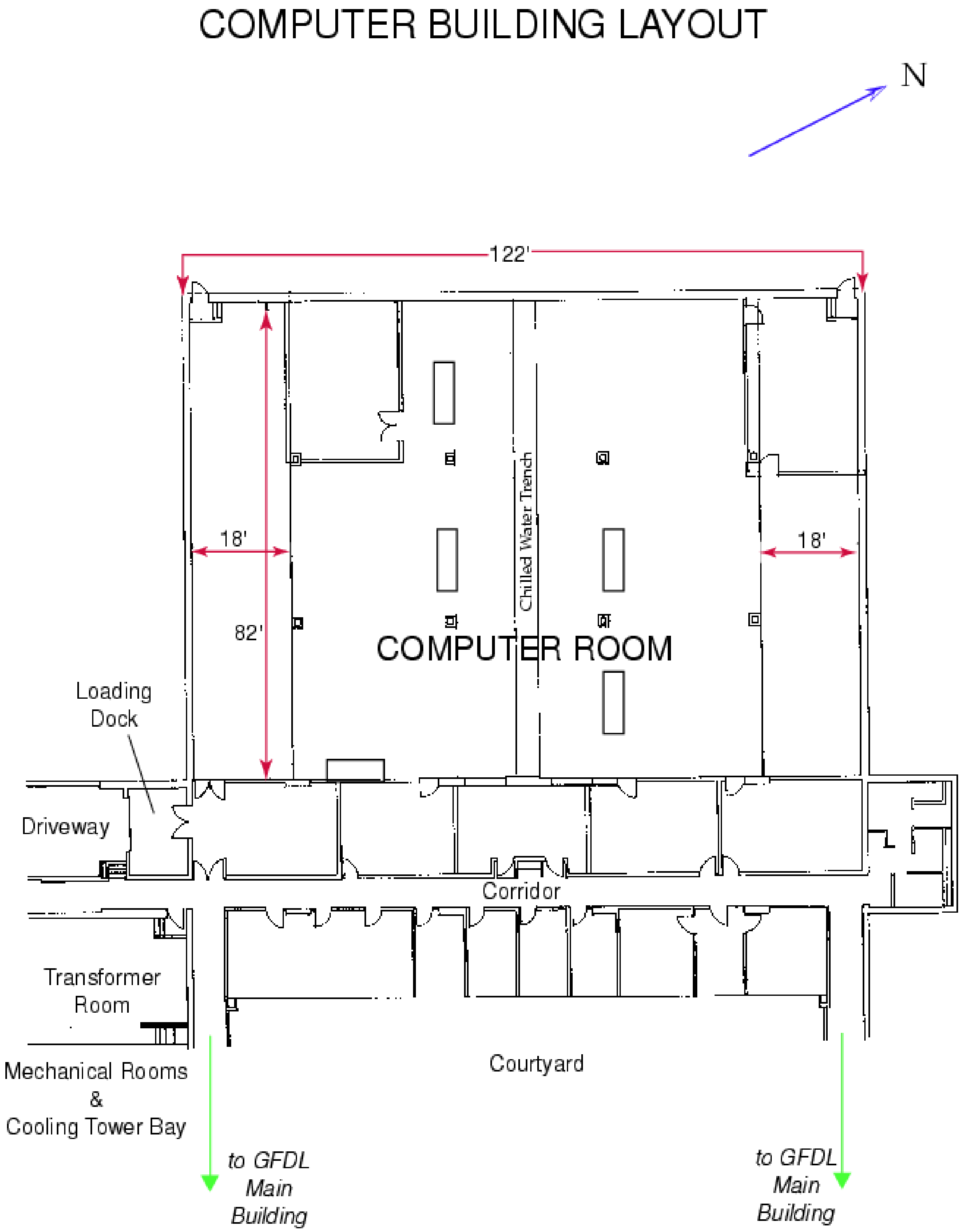


Figure 1 **Diagram of Computer Building**

demand is a total of these two numbers (595 kVA and 230 kVA), or 825 kVA. The system capacity is 1500 kVA. The additional capacity remaining in the Computer

Building's substation is the difference of the total capacity and the demand (1500 kVA minus 825 kVA), or 675 kVA. This number assumes that the PSE&G service feeders are sized for the full capacity of the substation and not the demand load. This should be verified with the utility company. Table I provides estimates of the power usage for current equipment in the Computer Room. The equipment listed is expected to be sharing power with the Initial System during the period of overlapping operation.

Table I. Estimated Power Usage of Current Equipment in the Computer Room

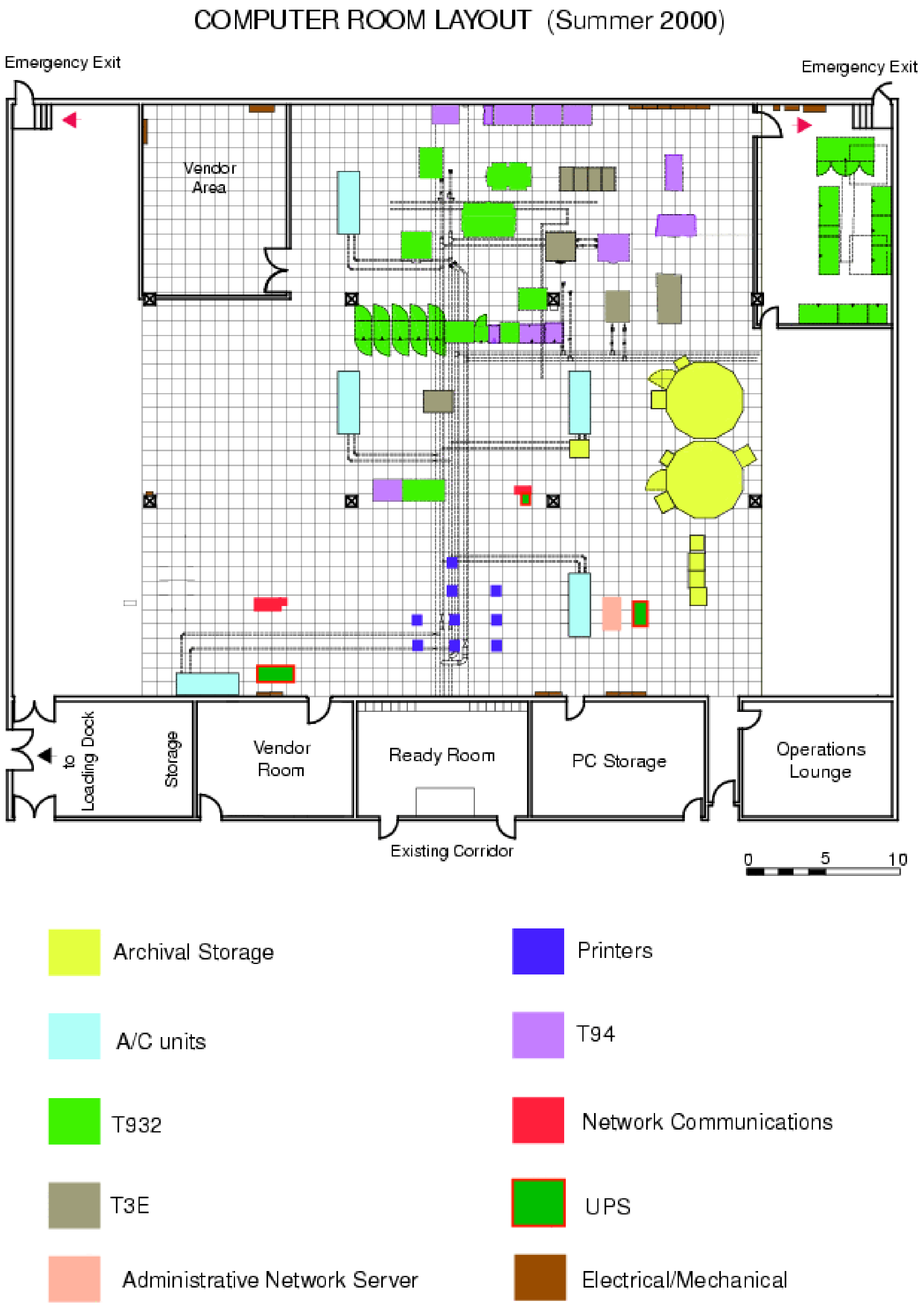
Equipment	Estimated Power Usage (KVA)
T932	410
T3E	77
T94	68
StorageTek Silos (2)	20
Air Handler A/C Units (7)	55
Printers (9)	12
Total	647

C.4.10.3 Available Cooling

Two centrifugal chillers and cooling towers make up the primary chilled water plant, which is located in the mechanical room and tower bay southeast of the Transformer Room (see Figure 1). These chillers are rated at 400 and 380 tons and are referred to as Chiller #2 and #3 respectively. The 400-ton chiller (Chiller #2) is being installed in the spring of 2000, along with new cooling towers and pumps. Chiller #3, installed in 1996, will be upgraded from 350 tons to 380tons capacity as part of this renovation. These systems are intended to be operated in such a way that only one chiller will be required on most days. The two chillers are intended to provide redundancy and to only be required on days in which cooling demands are unusually high. However, during the Initial System installation, both chillers may need to be operated on warm days in order to support both the new system and the SGI/Cray systems running in parallel. Chiller #4, located in the Transformer Room, is approximately 21 years old and will not be considered as a part of the normal operating chilled water plant after the spring 2000 renovation. This chiller may be phased out of operation after the renovation by the government or may be retained for emergency use, at the Government's option.

Cooling is delivered to the Computer Room through a six-inch piping system from the mechanical room at a temperature of 45 degrees Fahrenheit, plus or minus 2 degrees.

The pipe enters the computer room in a chilled water trench, 3-4 feet deep under the raised floor in the center of the Computer Room, as indicated in Figure 1. It is currently connected to five (5) air conditioning (blazer) units located on the raised floor, as well as to the refrigeration units of the existing SGI/Cray systems. Four (4) air conditioning units, each rated at 20 tons, are distributed in the center of the Computer Room, while one 40-ton unit is located near the doorway to loading dock. The locations of these five units is indicated by blue rectangles in Figure 2, which shows details of the room layout as of the summer of 2000. These existing air conditioning units, with compressors, refrigeration circuits, etc., are estimated to be 21 years old, and should be replaced if they are to be used as part of the site preparations because of their age. Two (2) 6,000 CFM chilled-water-cooled air handlers (not shown in Figure 2), with an estimated available capacity of 15 tons each, are mounted on the ceiling above the uninterruptable power supply (UPS) equipment for the T932, located in the room in the northern corner of the Computer Room. These two units appear to be in reasonable condition and may be reused for this area if new UPS equipment is placed in the same area as the present T932 UPS equipment. These units can only be used for this type of application per their present configuration.



<u>Figure 2</u> Equipment Layout in Computer Room (Summer 2000)

A chilled water air handler, located outside of the Computer Room, provides heating, cooling, and ventilation to the Computer Room for general non-equipment loads, and serves as the means for introducing outside air into this space. This unit also serves the offices, storage areas, corridors, and other spaces within the Computer Building. Figure 3 [to be provided later] shows the chiller configuration, with the three chiller units that have been cross-connected to allow for backup and parallel operation. This configuration will allow for up to 780 tons of cooling capacity with the two new chillers, Chiller #2 and #3, in operation at one time. Chiller #4 is not assumed to be used under normal circumstances. There is also a spare set of insulated, capped eight (8) inch lines that run between the mechanical room and the computer room trench. This set of 8" lines can be used to pipe up the Initial System equipment and the suggested new air conditioning equipment without disturbing the existing equipment.By accounting for the peak building cooling loads of the Main and Computer Buildings, the Government concludes that the maximum cooling capability that will be available to the Computer Room with both chillers (Chillers #2 and 3) operating is 5,100 KBTU/hr or 425 tons. This capacity would drop by approximately 175 tons to 250 tons or 3,000 KBTU/hr if the 400 ton chiller (Chiller #2) was shut down for service, and the 225-ton backup chiller (Chiller #4) was operated in its place. The use of Chiller #4 in this situation would also use power from the Computer Room substation as indicated in the previous section. Figure 3 does not include the present loads as detailed in the table below. The Initial System could initially utilize approximately 1092 KBTU/hr if Chillers #3 and 4 were operating with the existing SGI/Cray computers during the overlap period. The following estimates of cooling load by equipment category are provided as a guide. The equipment listed will be sharing cooling with the Initial System during the period of overlapping operation:

Table II. Estimated Cooling Load of Current Computer Room Equipment

Equipment	Estimated Cooling Load (BTU/hr)
T932	1,330,100
T3E and T94	450,000
StorageTek Silos (2)	55,000
Operators= Workstations	8,000
Printers (9)	34,000
Lighting	31,000
Total	1,908,100

C.4.10.4 Room Layout and Access

Figure 2 shows the computer room layout that is expected in the summer of 2000 prior to the beginning of site preparation. For purpose of reference, the front of the room as indicated in the following discussion refers to the bottom of the figure (nearest to the Ready Room), while the back of the room is at the top of the figure. The rooms shown at the bottom of the figure from left to right are:

- Loading Dock, which is designed to accept deliveries from 18-wheel trucks.
- Storage Room adjacent to Loading Dock, which also serves as a receiving/staging area for deliveries to the Laboratory.
- Contractor Room, which provides office space for contractor personnel of the existing system.
- Ready Room, where users interact with the operations staff and retrieve printouts from bins built into the wall between the Ready Room and the Computer Room.
- PC Storage Room, which currently provides space for miscellaneous storage.
- Door and Corridor, which is the primary entrance to the Computer Room.
- Operations Lounge.

Equipment access to the Computer Room from the Loading Dock is through two sets of double doors with clearances of 85 inches high by 70 inches wide. Two rooms have been constructed in the rear of the Computer Room. The room on the right contains the UPS equipment for the T932 system. The room on the left, constructed on the raised floor during a previous installation, is used by the current contractor as a storage and test area. Other details of the figure are as follows. The current SGI/Cray equipment is: T932 (green), which includes the UPS equipment located in the right rear room; T3E (gray); and T94 (purple), which includes UPS equipment located along the rear wall which supports both the T94 and T3E. The workstations associated with each of these systems are located in the center of the room. The two StorageTek silos and associated equipment are shown in yellow in the front right side of the room. Thin lines within the Computer Room indicate the locations of the current cooling pipes for the SGI/Cray equipment. The five (5) blue rectangles in the figure designate the positions of the air conditioning (blazer) units, as discussed previously. The X-symbols within boxes indicate the locations of the eight support columns in the middle of the room. Additional equipment expected to be in the room in summer of 2000 are: nine (9) Government-owned printers (2 HP LaserJet 8100N, 4 HP LaserJet 5M, and 3 Tektronix Phaser 740 printers) (dark blue), located in the front of the Ready Room; network communications equipment (bright red); Network UPS equipment (dark green with red border); the administrative network server (pink); and electrical panels (brown).

C.4.10.5 Available Floor Space

As indicated in Figure 1 and 2, the entire Computer Room is 8,405 square feet in size, with dimensions of 102.5 feet by 82 feet. The raised floor area, indicated by a grid of squares depicting each two-by-two-foot floor tile, is 7,052 square feet in size, with dimensions of 86 feet by 82 feet. The raised area of the Computer Room is 24 inches above the concrete subfloor.

The existing raised floor sections were installed at three different times. The locations of these sections are indicated by Figure 4, which distinguishes these three groups by color according to the year in which each was installed. The blue and red sections of Figure 4, which were installed respectively in 1990 and 1995, have anti-static carpeted panels. The raised floor is electrically interconnected to provide a common electric reference. The raised floor is designed to support a uniform live load of 250 pounds per square foot, with a deflection of not more than 0.040 inch. Great care must obviously be taken in moving heavy equipment across any raised floor so as to distribute equipment loads evenly. The raised floor shown in green in Figure 4 was installed in 1980. Any sections of this older flooring that will be used to support new equipment shall be replaced as part of the site preparations before new equipment is moved onto it. In fact, it is recommended that the contractor test the integrity of floor sections and supports and, if necessary, replace any floor sections and/or supports that are inadequate before installing new equipment. The Government=s past strategy for floor space usage in the GFDL Computer Room has been to limit the amount of space available to the new contractor to no more than half of the total raised floor space within the room. The purpose of this was to leave sufficient space unoccupied so that the follow-on contractor would be able to install and operate the next system in parallel with the current system. With this objective in mind, the Government considers it to be desirable that contractors restrict their use of floor space, both raised and solid, to no more than half of the total space within the Computer Room. If the new equipment uses more than half of the available floor space, the proposal should provide recommendations on how the Government can design the follow-on procurement and installation in order to provide for overlap of systems.

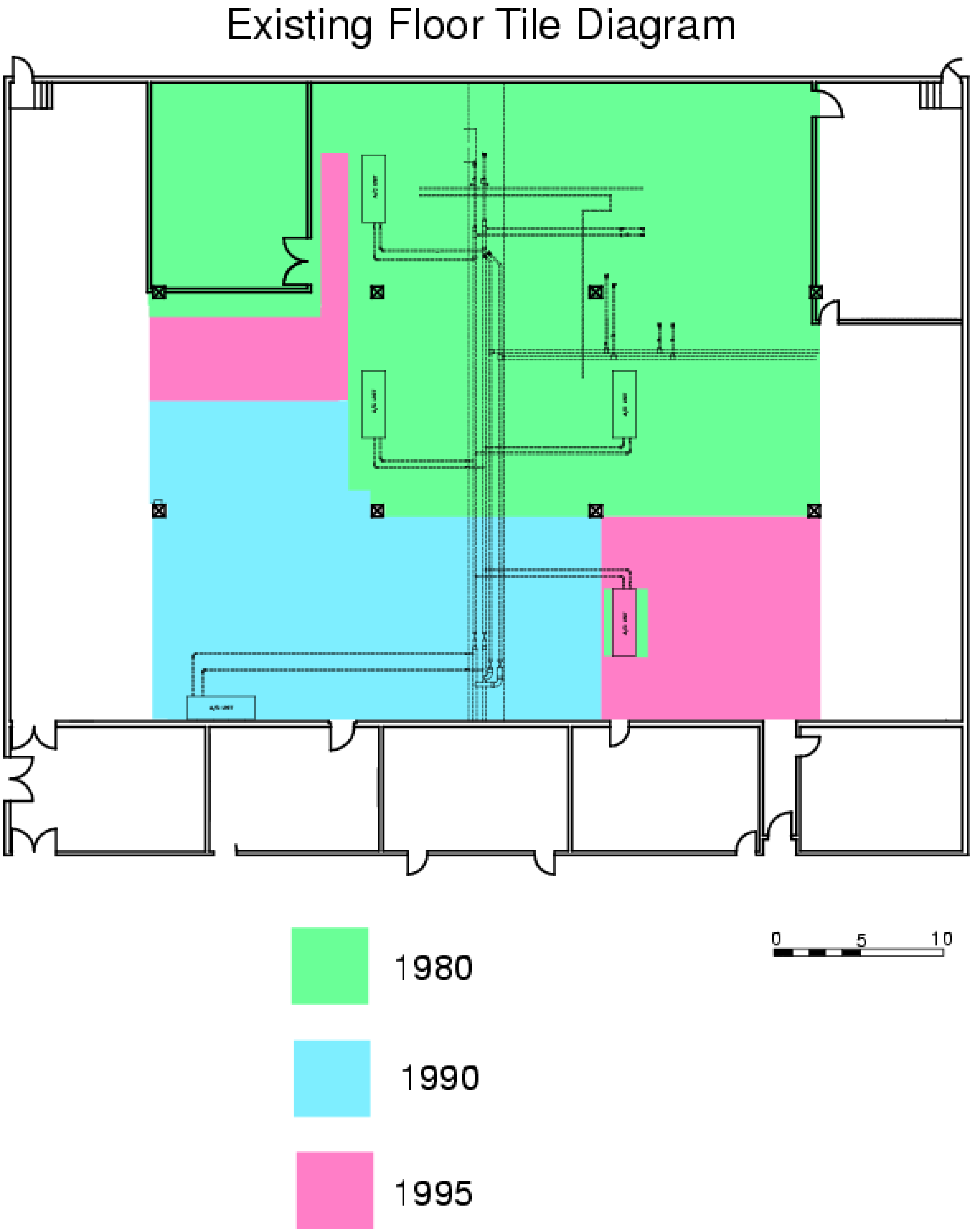
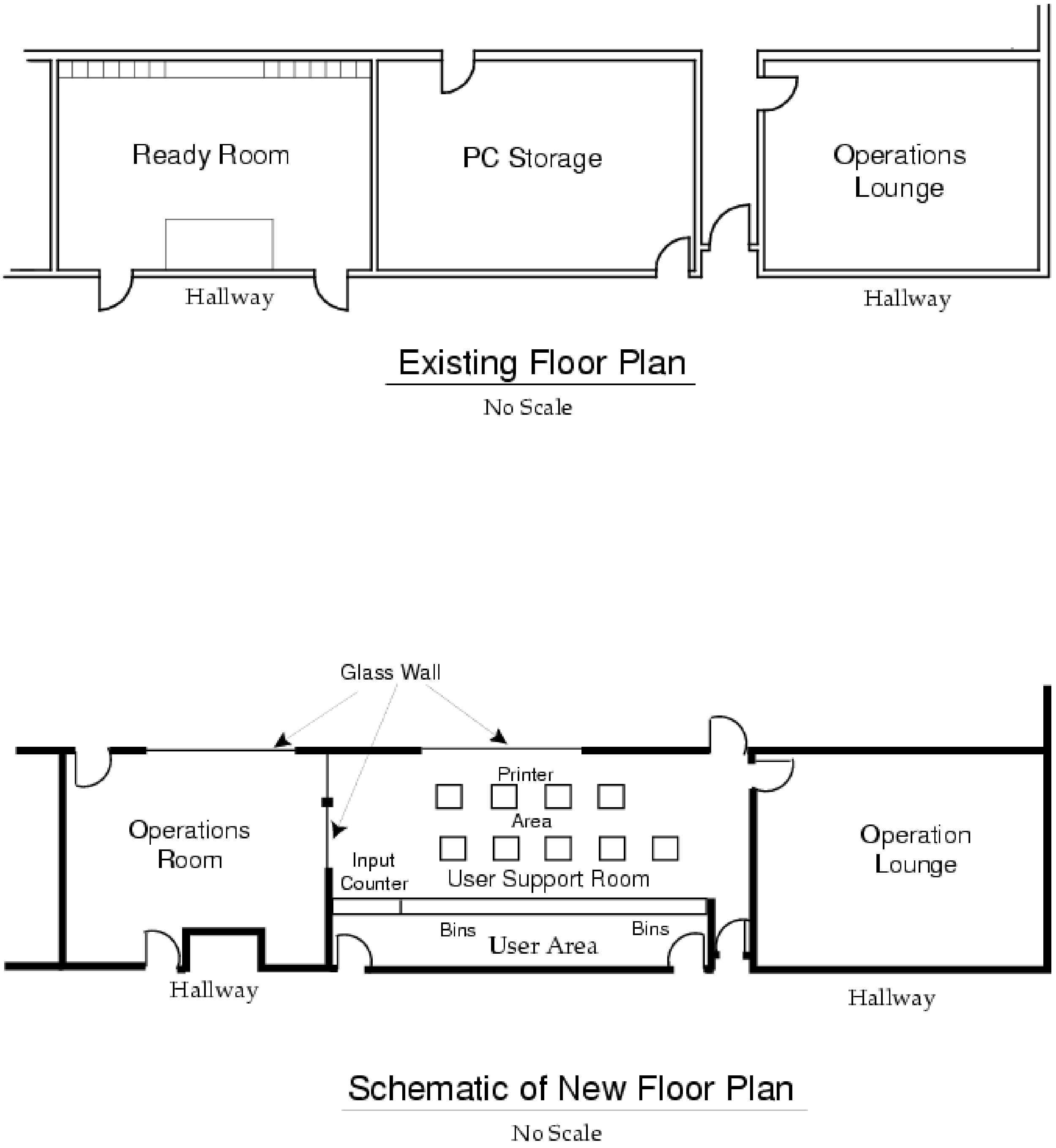


Figure 4. **Location of Raised Floor Sections According to Year of Installation**

C.4.10.6 Facility Renovations to Provide Rooms for Operators and Printers

Under the current arrangements, the GFDL Operations staff operates the SGI/Cray systems from workstations located in the Computer Room itself. In addition, the operations staff is responsible for managing printers located in the Computer Room, where these printers are the primary means by which GFDL users produce printed output. The Government has concluded that the operations control area should be moved out of the main computer room for two reasons: to provide the Operations staff with a quiet work environment, and to increase the raised floor space available for equipment. Users will need to have reasonable access to the Operations staff, while maintaining acceptable physical security. In addition, user bins should be accessible by users while being located close to the printers, which are also moved out of the main computer room. With these objectives in mind, the Government requires that the current Ready Room and PC Storage Room (Figure 2) be renovated to provide an Operations Room and a User Support Room. The Operations staff will oversee and manage the systems and networks from the Operations Room. The User Support Room will be divided into a Printer Area and a User Area, separated by a wall containing user bins and an input counter. The user bins are cubicles reserved for individual users' printed output. The input counter is an open counter area where users can communicate directly with the operations staff and packages can be received by Operations.The upper frame of Figure 5 indicates an enlargement of the current layout of the Ready Room, PC Storage Room, and Operators= Lounge as taken from Figure 2. The lower frame of this figure shows a schematic of a proposed layout for the Operations Room and User Support Room and their position relative to the Operators' Lounge.



Existing Floor Plan

No Scale

Schematic of New Floor Plan

No Scale

Figure 5. **Original Floor Plan and Schematic of Proposed New Design**

Both the Operations Room and the Printer Area will have at least a 6-inch raised floor to accommodate wiring for the various consoles and printers. The back walls of the Operations Room and Printer Area (adjacent to the Computer Room) will have large glass windows to provide a view of the Computer Room. All doors leading to the Computer Room shall be equipped with DOC-approved security devices. Presently, the Laboratory has card readers and cipher locks installed on doors accessing the Computer Room.

The design should provide for quick access to the fire alarm panel and environmental monitoring system. All necessary operating consoles, including consoles for network servers, shall be relocated to the Operations Room and shall be designed to provide convenient access. Cameras shall be placed in strategic locations within the Computer Room with monitors installed in the Operations Room. This will enable the Operations staff to observe the Computer Room in areas that are not visible through the glass windows.

This construction is likely to require modifications to mechanical and electrical systems in order to provide the additional capacity in these rooms for the equipment being relocated. Additional ventilation may also be needed.

C.5 Options

Mandatory options for minimums of 512MB and 1GB of main memory per processor on the initial LSC shall be offered.

Three one-year options that provide guaranteed increased performance levels on both the LSC and AC shall be offered. Only the performance levels proposed for these option years will be considered in awarding the HPCS contract.

C.6 Definitions

The following definitions, listed in alphabetical order, will be used in this Statement of Need:

Aggregate Tape Positioning Rate. The aggregate tape positioning rate P = N * 3600 / (L + R + S), where N = number of drives proposed for small frequently-accessed files, L = the load and thread time in seconds, S = the median search time for the proposed tape volume in seconds, and R = the median rewind time for the proposed tape volume in seconds.

Availability Level. The availability level of a computer is a percentage figure determined by dividing the operational use time by the difference between wallclock and null time.

Byte. Eight (8) bits.

Cluster. A collection of nodes with a dedicated interconnect.

Computer. The maximum set of nodes that may be unavailable during the repair of any subset of those nodes.

DCE. Distributed Computing Environment.

Decimal Gigabyte (dGB). Ten (10) to the power of nine (9) bytes.

Decimal Gigaword (dGW). Ten (10) to the power of nine (9) words.

Decimal Kilobyte (dKB). Ten (10) to the power of three (3) bytes.

Decimal Kiloword (dKW). Ten (10) to the power of three (3) words.

Decimal Megabyte (dMB). Ten (10) to the power of six (6) bytes.

Decimal Megaword (dMW). Ten (10) to the power of six (6) words.

Decimal Terabyte (dTB). Ten (10) to the power of twelve (12) bytes.

Decimal Teraword (dTW). Ten (10) to the power of twelve (12) words.

DFS. Distributed File Service.

DNS. Domain Name System.

Failover. In the event of a failure, resources are available to assume and resume the tasks that were using the failed resources without user or operator intervention.

Gigabyte (GB). Two (2) to the power of thirty (30) bytes.

Gigaword (GW). Two (2) to the power of thirty (30) words.

Kilobyte (KB). Two (2) to the power of ten (10) bytes.

Kiloword (KW). Two (2) to the power of ten (10) words.

Megabit (Mb). Two (2) to the power of twenty (20) bits.

Megabyte (MB). Two (2) to the power of twenty (20) bytes.

Megaword (MW). Two (2) to the power of twenty (20) words.

Migration. The movement of the contents of a disk-resident file to tape volume(s), or the movement of a file's contents from tape volume(s) back to disk storage.

NFS. The Network File System as defined by specifications placed into the public domain by Sun Microsystems, Inc.

NIS. The Network Information Service as defined by specifications placed into the public domain by Sun Microsystems, Inc.

Node. A collection of one or more processors and one or more memory modules that communicate among themselves with a different communication architecture (in protocol, organization, or performance) than they use to communicate with other nodes.

Null Time. The time during which equipment is unavailable to the government, exclusive of charged downtime.

Operational Use Time. The time during which equipment is available to the government, exclusive of preventive maintenance time, remedial maintenance time, standby time, or Contractor-caused machine failure. Partial credit will be given for equipment operating in degraded mode (for example, when a portion of the processors, memory, disk, etc. on a computer is unavailable). The government may declare the entire HPCS down even if parts of the HPCS are available.

Preventive Maintenance (PM). That maintenance performed by the Contractor which is designed to keep the equipment in proper operating condition. It is performed on a scheduled basis.

Primetime. 7am to 7pm Eastern Time, Monday through Friday.

Processor. The minimum physical or functional unit on which a unitasked job can run.

Remedial Maintenance (RM). That maintenance performed by the Contractor which results from Contractor-supplied equipment or operating software failure. It is performed as required and therefore on an unscheduled basis.

Terabyte (TB). Two (2) to the power of forty (40) bytes.

Teraword (TW). Two (2) to the power of forty (40) words.

Word. Sixty-four (64) bits.